

Forecasting stock prices using machine learning technique

Abstract:

the stock market is an emerging sector in any country of the world . Many people directly related to this sector . Along with the development with the stock market ,forecasting become an important topic .Since finance market has become more and more competitive, stock price prediction has been a hot research topic in the past few decades .predicting stock price is regarded a challenging task because stock market is essentially non linear ,non-parametric,noisy,and a chaotic system .Trend of a market depends on many things like liquid ,money human behavior, news related to stock market etc. All this together controls the behavior of trends in a stock market with the advancement of the computing technology we use machine learning technique,like Support Vector Regression,K-nearest-neighbor,liner Regression for analyzing time series data to predict stock price. In this paper I try to develop a forecasting model with multiple method to find the best forecast of the stock price.

‘Keywords— Google stock, , Time Series data, SVR, KNN-Regressor, liner regression’

introduction:

The goal is to take time series data, find the equation that best fits the data, and be able forecast out a specific value. Time series data is a continuous data statistical observations recorded over a specific period of time. This model will try to understand the pattern of the continuous data with different method and produce a best fit line that fits the data. The target is to determine the future stock price and improve their strategy for future

This paper provides a hybrid methodology that combines both support vector machine ,liner regression and knn regression technique for dealing with high dimensional feature ,finding relation among the data and create a best fit line which will give us a significant accuracy and representational advantages

linear regression is a linear approach to modelling the relationship between a scalar response dependent variable and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called **multiple linear regression**

support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model by creating a hyper plane that assigns new examples to one category or the other . in support vector regression (SVR). The model produced by support vector classification (as described above) depends only on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin.

In pattern recognition, the ***k*-nearest neighbors algorithm (*k*-NN)** is a non-parametric method used for classification and regression. In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression:

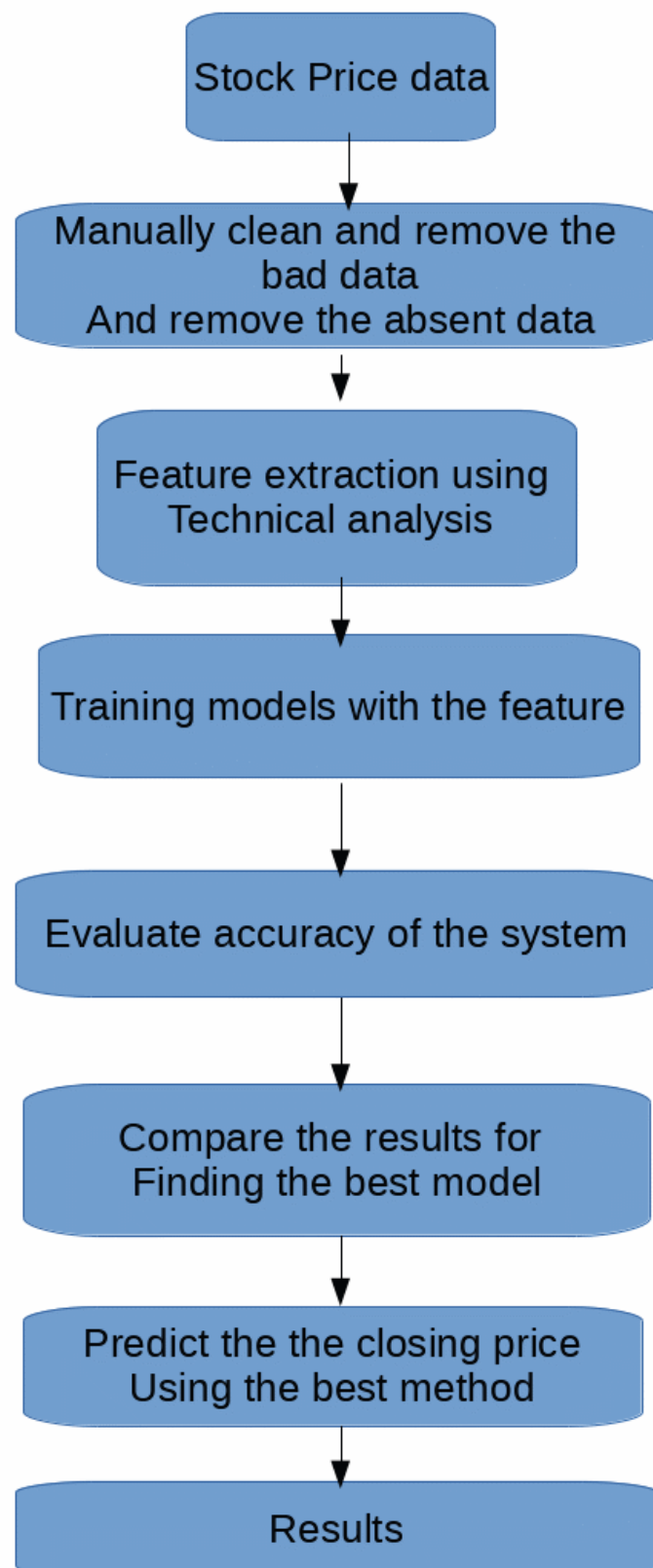
in case of the knn regression the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors. a useful technique can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones

ARCHITECTURE OVERVIEW

The multiple data source prediction system has a two tier architecture. The top tier is dedicated to preparing the datasets from multiple information sources to make them ready for the predication tasks in the next tier. It is composed of two major parts. The first part is data prepossessing . In this process we process the data by adding more feature and removing unnecessary feature and removing the bad data and also the absence of the data . The second part is the data alignment. The second tier is dedicated to the market volatility analysis and prediction through the model integration and training, which uses multiple kernel learning methodology to train the classifier. It consists of three tasks: First, we build one regression model per source. Second, we train the model with the same datasets ,and find the accuracy by cross validation of the result and finally we perform the comarison between the those result of different model to find the best prediction model In this paper, we use the multi-kernel learning method .

A. Data Preprocessing

filter out other unimportant feature from the feature because not all the feature will be included into the final feature list. the reason behind it is the unnecessary feature and those value which has no relation with the stock market prediction will reduce the accuracy of the prediction



METHODOLOGY

Different kind of machine learning models have been developed for forecasting of time series data. In this section we describes how our forecasting model's algorithm work

data processing: in order to do apply the into our model we have to first process the data. Creating appropriate feature for making the perfect and most important data to analyze with to predict the future data

Liner Regression is used for creating a a best fitting line with the help of the previous data point

SVM is used to deal with the high dimensional feature and create the best hyper plane that fits the data

k -nearest neighbors finds find the average of the values of its k nearest neighbor and give a average fit line

then after the comparison of the accuracy and based the accuracy we visualize the data with the best fit line.